

Санкт–Петербургский государственный университет

Балавнев Данил Андреевич

Выпускная квалификационная работа

Использование нейронных сетей в задаче прогнозирования
закупок товаров

Уровень образования: бакалавриат

Направление 02.03.02 «Фундаментальная информатика и
информационные технологии»

Основная образовательная программа СВ.5003.2016 «Программирование
и информационные технологии»

Профиль «Автоматизация научных вычислений»

Научный руководитель:

Кандидат физ.-мат. наук, доцент кафедры
теории систем управления электрофизической аппаратурой
Козынченко Владимир Александрович

Рецензент:

Доктор физико-математических наук,
профессор кафедры теории управления
Котина Елена Дмитриевна

Санкт-Петербург

2020 г.

Содержание

Введение	3
Постановка задачи.....	5
Глава 1. Существующие решения.....	6
1.1 Excel.....	6
1.2. Novo Forecast Pro.....	7
1.3 Forecast NOW.....	9
Глава 2. Разработка приложения.....	11
2.1 Разработка нейронной сети.....	13
2.1.1 Выделение признаков	14
2.1.2 Архитектура нейронной сети	16
2.1.3 Реализация нейронной сети.....	22
2.2 Разработка веб-сервиса	27
2.2.1 Реализация веб-сервиса	29
Глава 3. Анализ полученного решения	32
Заключение	34
Список литературы	35

Введение

В условиях нестабильного рынка деятельность любого бизнеса зависит от точности определения необходимого объема закупок. Обычно из всего многообразия товара, представленного компанией, ходовыми можно считать лишь четверть, остальное, как правило, раскупается с меньшей интенсивностью. Наиболее распространенной является условная классификация товара на две категории: скоропортящиеся и длительного хранения. К первой группе преимущественно относятся продукты питания, ко второй различные товары хозяйственного назначения. Кроме этого весь ассортимент продукции также можно разделить на сезонные товары и товары имеющие постоянный спрос. Например к сезонным товарам можно отнести сувенирную предпраздничную продукцию, спрос на которую в остальные месяцы будет крайне мал.

На плечи закупщика ложится тяжелая задача расчета оптимального количества ликвидного товара. Низкий товарный запас порождает дефицит и негативно сказывается на лояльности клиентов, и напротив, чрезмерное количество влечет издержки хранения.

Для определения закономерностей в данных по продажам и построения прогноза, учитывающего колебания спроса и сезонности необходимого количества продукта, отлично подходят нейронные сети. Приложение, использующее нейронные сети для расчета необходимого количества товара, позволит не только оптимизировать закупки, но и предсказать покупательский спрос, что, в свою очередь,

поможет определить поведение потенциальных потребителей и увеличить прибыль предприятия с минимальными затратами.

В данной работе будут рассмотрены уже существующие программные решения, предоставляющие инструменты для прогнозирования закупок и покупательского спроса, выявлены их основные достоинства и недостатки, а также предложено альтернативное решение.

Постановка задачи

Целью данной работы является создание приложения для прогнозирования закупок товаров, которое позволяло бы на основе предоставленных данных производить анализ и строить прогноз на количество товара, необходимого для приобретения. Приложение должно удовлетворять следующим критериям:

1) Возможность составлять как краткосрочные, так и среднесрочные прогнозы.

2) При составлении прогноза должна учитываться сезонность продаж.

3) Интерфейс для работы с приложением должен быть интуитивно понятен и удобен в работе для закупщика. Результаты работы должны демонстрироваться в наглядном виде.

4) Нейронная сеть должна быть устойчива к выбросам данных.

5) Приложение должно выдавать высокую точность прогноза.

Таким образом, в задачи данной работы входит:

- Анализ существующих популярных решений, используемых для прогнозирования объема необходимого количества товара.
- Разработка, обучение и тестирование нейронной сети.
- Разработка приложения.
- Анализ полученного решения.

Глава 1. Существующие решения

В первую очередь, необходимо рассмотреть уже существующие решения, используемые закупщиками для расчета необходимого количества товара.

1.1 Excel

Microsoft Office Excel – программа, предназначенная для работы с электронными таблицами, позволяющая хранить и анализировать информацию. Данная программа предоставляет различные инструменты для проведения статистических расчетов.

Для работы с данными предприниматели создают шаблонную таблицу, в которую вносят все объемы своих продаж. Для составления прогноза используются встроенные функции, такие как[1]:

- Функция TREND. Рассчитывает значения в соответствии с линейным трендом, через заданный набор зависимых значений. Аппроксимирует прямой линией массивы объема продаж и порядковые номера месяцев, используя метод наименьших квадратов.
- Функция FORECAST. Вычисляет будущее значение используя метод линейной регрессии. Разница между TREND и FORECAST заключается в том, что функция FORECAST может прогнозировать будущие значения только на основе существующих значений, а функция TREND может рассчитывать как текущие, так и будущие тренды. Функция FORECAST используется в качестве обычной формулы и возвращает одно новое значение y для одного значения x. Функция TREND

используется в качестве формулы массива и вычисляет несколько значений y для нескольких значений x .

- Функция LINEST. Вычисляет статистику для ряда, используя метод наименьших квадратов, чтобы вычислить прямую линию, которая наилучшим образом аппроксимирует данные.

Microsoft Excel требует много ручной работы, чтобы собрать все данные, экспортировать и упорядочить их в электронной таблице. У менеджера, который занимается ведением учета и прогнозированием, есть множество задач, которые необходимо выполнить, а именно: выбрать наилучшую модель, установить сроки расчета, запустить модель прогнозирования, проанализировать данные. Расчеты и сбор данных усложняются, когда необходимо проанализировать сезонные продажи. Учитывая факт вмешательства человека и ручного выполнения большого количества операций, риск совершения ошибок повышается.

1.2. Novo Forecast Pro

Novo Forecast Pro - приложение, предназначенное для построения прогнозов на основании больших объёмов данных в Excel, которое анализирует характеристики каждого временного ряда и автоматически подбирает наиболее точную модель прогнозирования. Novo Forecast Pro включает в себя такие модели как [1]:

- Модели скользящего среднего. Применяется в случае сильно изменчивых коротких данных.
- Модели экспоненциального сглаживания.

- Сезонные модели ARIMA и модель Бокса-Дженкинса. Применяется для стабильных данных.
- Модели кривых роста.
- Модель прерывистого спроса Кростона и дискретные модели спроса применяются в случае разреженных и недостаточных данных.
- Модели динамической регрессии.

Выбор модели для прогноза осуществляется с использованием нескольких показателей точности, а именно: MAPE, MAD, GMRAE

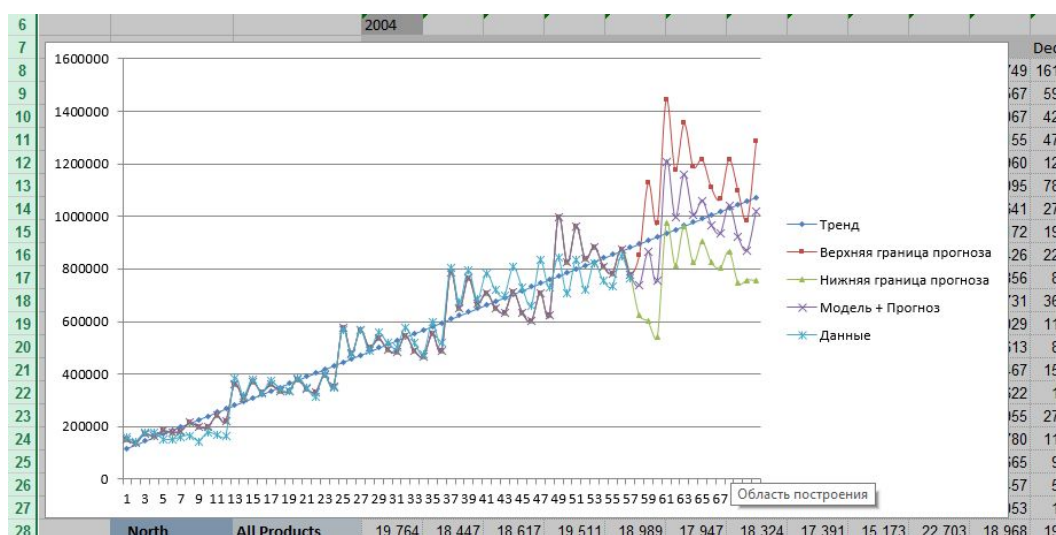


Рис.1: Пример работы Novo Forecast Pro

К преимуществам данного решения можно отнести:

- Расчет производится как по полным временным рядам, так и по неполным.
- Анализ характеристик каждого временного ряда.
- Расчет точности и ошибок прогноза.

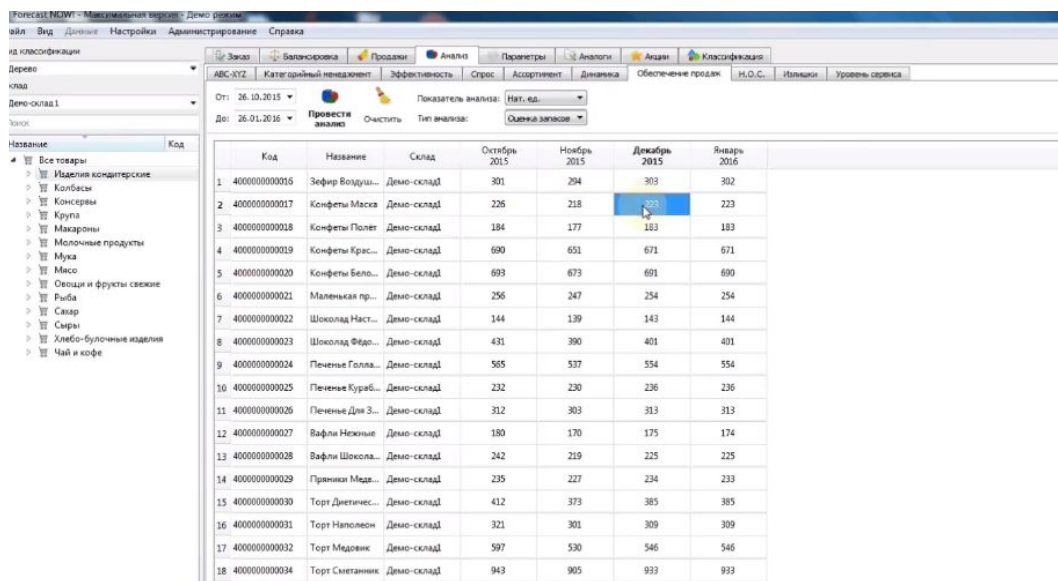
- Графический анализ отклонений плана от прогноза.

Недостатками можно считать:

- Привязка работы приложения к Excel. Для расчета все данные необходимо экспортировать в электронную таблицу.
- Все данные должны быть приведены к определенной структуре.

1.3 Forecast NOW

Forecast NOW - система управления складскими запасами, предназначенная для розничных и оптовых торговых предприятий. Из пользовательской системы автоматизации передаются данные программе, которая строит прогноз, на основании внешних факторов и поданных ограничений вычисляется оптимальный товарный запас и формируется отчет о необходимых закупках.



The screenshot shows the 'Forecast NOW' application window. On the left is a tree view of goods categories. The main area displays a table with columns for 'Код' (Code), 'Название' (Name), 'Склад' (Warehouse), and forecasted stock levels for 'Октябрь 2015', 'Ноябрь 2015', 'Декабрь 2015', and 'Январь 2016'. The table lists 18 different goods, including various types of candies, biscuits, and cakes.

Код	Название	Склад	Октябрь 2015	Ноябрь 2015	Декабрь 2015	Январь 2016
1 40000000000015	Зефир Воздуш...	Демо-склад	301	294	303	302
2 40000000000017	Конфеты Масса	Демо-склад	226	218	223	223
3 40000000000018	Конфеты Полет	Демо-склад	184	177	183	183
4 40000000000019	Конфеты Крас...	Демо-склад	690	651	671	671
5 40000000000020	Конфеты Бело...	Демо-склад	693	673	691	690
6 40000000000021	Маленькая пр...	Демо-склад	256	247	254	254
7 40000000000022	Шоколад Наст...	Демо-склад	144	139	143	144
8 40000000000023	Шоколад Фидо...	Демо-склад	431	390	401	401
9 40000000000024	Печенье Голла...	Демо-склад	565	537	554	554
10 40000000000025	Печенье Кураб...	Демо-склад	232	230	236	236
11 40000000000026	Печенье Дня З...	Демо-склад	312	303	313	313
12 40000000000027	Вафли Нежные	Демо-склад	180	170	175	174
13 40000000000028	Вафли Шокола...	Демо-склад	242	219	225	225
14 40000000000029	Пирожки Мед...	Демо-склад	235	227	234	233
15 40000000000030	Торт Диетичес...	Демо-склад	412	373	385	385
16 40000000000031	Торт Наполеон	Демо-склад	321	301	309	309
17 40000000000032	Торт Медовик	Демо-склад	597	530	546	546
18 40000000000034	Торт Сметанник	Демо-склад	943	905	933	933

Рис. 2: Пример работы Forecast Now

Программа использует нейронные сети, причем автоматический выбор структуры сети и формата входных данных осуществляется на основе анализа динамики продаж, горизонта прогнозирования. Для работы с разреженным спросом, который имеет место, например, в сфере торговли автозапчастями и стройматериалами, используется модель прогнозирования редких продаж.

Программа предоставляет возможность учета дополнительных факторов, например, курс доллара, погода, изменение цены товара, складские остатки.

Преимуществами данного приложения являются:

- Визуализация данных продаж, остатков, цен, прибыли и прогнозов спроса по товарам и товарным группам.
- Возможность построения как краткосрочного (до 4 недель), так и долгосрочного прогноза (до 6 месяцев).
- Применение фильтров по свойствам товаров позволяет эффективно производить анализы и строить прогнозы для различных целей.
- При анализе учитываются остатки запасов.

В качестве недостатков можно выделить:

- Forecast Now не подходит для ведения и учета малого бизнеса, онлайн магазинов без систем автоматизации.
- Прогнозирование временных рядов только с использованием нейронных сетей представляется недостаточным. В частности, при малых объемах данных нейронные сети в сравнении с

простой линейной моделью, могут давать прогнозы более низкого качества. [2]

- Платная лицензия на использование, цена которой начинается от 740 000 р.

Глава 2. Разработка приложения

Существует достаточно большое количество методов, которые можно использовать для прогнозирования необходимого количества товара. Выделим следующие:

- Прогнозирование с помощью метода скользящей средней. Данный метод вычисляет общую тенденцию в наборе данных, заменяя в простейшем случае значения средним арифметическим интервала[3]. Скользящая средняя не учитывает “возраст” наблюдения, для учета влияния прошлых наблюдений используется прогнозирование с помощью скользящей средней, в которой веса подчиняются экспоненциальному закону. Экспоненциальное скользящее среднее является характеристикой последних значений ряда динамики, которым присваивается наибольший вес. Идея этого подхода состоит в том, чтобы давать различный вес каждому периоду внутри выбранного интервала.
- Прогнозирование на основе колебаний. [4] Под колебаниями понимаются изменения уровня динамического ряда, вызванные сменой временем года. Сезонные изменения строго цикличны. Метод состоит из трех этапов:

1. За продолжительный отрезок времени исследуется тенденция изменения прогнозируемого показателя.
 2. Анализируются сезонные изменения прогнозируемого показателя за короткий промежуток времени и строится график “сезонной волны”.
 3. Прогнозируется динамика изменения показателя в поквартальном либо помесечном разрезе.
- Прогнозирование с помощью регрессионных моделей (линейные и нелинейные, авторегрессионные). В простой линейной регрессии прогнозируется одна зависимая переменная (например, количество необходимого товара) и одна независимая переменная[5]. Значения независимой переменной обычно определяют значения зависимой переменной. В модели авторегрессии мы прогнозируем следующий шаг, используя линейную комбинацию наблюдений за предыдущие шаги.

У каждого из приведенных методов есть как достоинства, так и недостатки, например, расчет на основе колебаний сезонности пригоден для товаров с явно выраженной сезонной составляющей, а линейная регрессия, напротив, не учитывает сезонные колебания. Методы группы скользящих средних используют в своей основе простой алгоритм, который дает достаточную эффективность при наличии ярко выраженного долгосрочного (возрастающего, убывающего) тренда [6]. Слабым местом данного метода, является субъективность выбора периода усреднения скользящей средней, который может находиться в больших пределах, что приведет к снижению качества прогноза.

Центральным моментом в анализе временных рядов является то, что существует множество непредсказуемых факторов, приводящих к сложным временным зависимостям. Нейронные сети являются одним из методов, который способен учесть большое количество факторов, сезонность и историю продаж. Они устойчивы в разумных пределах к помехам во входных данных и могут даже поддерживать обучение и прогнозирование при наличии пропущенных значений. Также важным свойством нейронной сети является обобщающая способность - способность выдавать правильные результаты не только для данных участвующих в обучении, но и для новых. Нейронные сети не исходят из априорных предположений о функциональной взаимосвязи ряда и его независимых переменных, что является весьма желательной характеристикой в случае работы с временными рядами.

2.1 Разработка нейронной сети

При определении объема товара для закупки можно отталкиваться от трех временных рядов: на основании данных о продажах, на основании остатков товара на складе, редко – на основании закупок. Возможны комбинационные варианты, а также добавление спроса. Логичнее всего отталкиваться от данных по продажам.

В качестве датасета для разработки и тестирования нейронной сети была взята двухгодичная база данных продаж онлайн магазина содержащая в себе 1 400 000 записей. Компания специализируется на продаже сувенирной продукции.



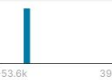
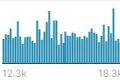
Invoice	StockCode	Description	# Quantity	InvoiceDate	# Price	# Customer ID	Country
Invoice number, Nominal. A 6-digit integral number uniquely assigned to each transaction. If this code starts with the letter 'c', it	Product (item) code. Nominal. A 5-digit integral number uniquely assigned to each distinct product.	Product (item) name. Nominal.	The quantities of each product (item) per transaction. Numeric.	Invoice date and time. Numeric. The day and time when a transaction was generated.	Unit price, Numeric. Product price per unit in sterling (£).	Customer number. Nominal. A 5-digit integral number uniquely assigned to each customer.	Country name, Nominal. The name of the country where a customer resides.
53628 unique values	5305 unique values	5699 unique values					United Kingdom 92% EIRE 2% Other (68175) 6%
489434	85848	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
489434	22841	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.1	13085.0	United Kingdom
489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
489434	22864	PINK DOUGHNUT TRINKET POT	24	2009-12-01 07:45:00	1.65	13085.0	United Kingdom
489434	21871	SAVE THE PLANET MUG	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
489434	21523	FANCY FONT HOME SWEET HOME DOORMAT	10	2009-12-01 07:45:00	5.95	13085.0	United Kingdom
489435	22350	CAT BOWL	12	2009-12-01 07:46:00	2.55	13085.0	United Kingdom
489435	22349	DOG BOWL , CHASING BALL DESIGN	12	2009-12-01 07:46:00	3.75	13085.0	United Kingdom
489435	22195	HEART MEASURING SPOONS LARGE	24	2009-12-01 07:46:00	1.65	13085.0	United Kingdom

Рис. 3: Пример данных

2.1.1 Выделение признаков

Для повышения качества работы нейронной сети проведем feature engineering. Проанализируем данные, посчитаем за каждый час сумму всех покупок.

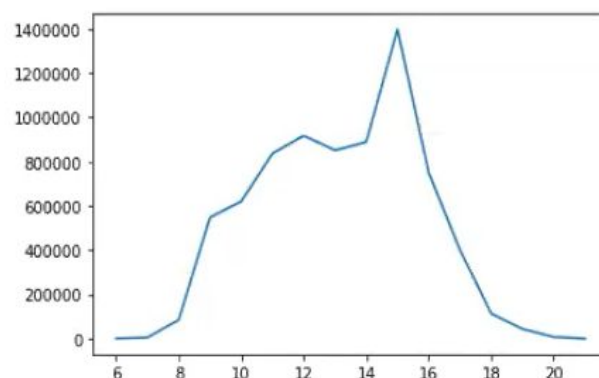


Рис. 4: Распределение объема продаж за сутки

Как мы видим, пик продаж достигается в период с 15 до 16 часов, следовательно, в качестве признака целесообразно учитывать

время совершения продажи. Другими свойствами, от которых может зависеть количество проданного товара являются:

- Был ли в этот день праздник.
- Количество дней до ближайшего праздника. В предпраздничные дни возможно сильное повышение уровня продаж и, если запасы окажутся недостаточными, это чревато высоким уровнем упущенной прибыли.
- Общий тренд. Изменение средней цены товара в зависимости от времени, цена постепенно повышается или понижается.

date	Quantity	is_holiday	lr_coef_1	lr_coef_2	nearest_hday_timedelta
2009-12-01	22288	0	0.003636	15.985310	-1.565709e+09
2009-12-02	27345	0	0.006317	17.402811	-2.563935e+09
2009-12-03	46502	0	0.002621	17.217994	-3.601484e+09
2009-12-04	19768	0	0.001425	16.744797	-4.524210e+09
2009-12-05	4722	0	-0.002835	21.122722	-5.570877e+09

Рис. 5: Вектор признаков

Таким образом входной вектор будет состоять из 5 признаков:

- Quantity - количество проданных товаров за день.
- is_holiday - был ли в этот день праздник.
- lr_coef_1, lr_coef_2 - коэффициенты линейного полинома полученные при аппроксимации количества проданных товаров ко времени в течение которого он продавался.
- nearest_hday_timedelta - количество дней до ближайшего праздника.

```

dd = {}
try:
    for sc in df['StockCode'].unique():
        mask = df['StockCode'] == sc
        a, b = np.polyfit(df[mask]['time_from_start'].unique(), df[mask].groupby('time_from_start')['Quantity'].agg('sum'), 1)
        dd[sc] = a, b
    %telemetry 'lr coefs are done'
except BaseException as e:
    %telemetry 'Exception: ' + str(e)

```

Рис. 6: Вычисление тренда

2.1.2 Архитектура нейронной сети

При использовании сетей прямого распространения подразумевается, что все входы и выходы независимы, но для задачи прогнозирования необходимого количества закупок это не подходит. Рекуррентные нейронные сети (RNN), называются рекуррентными, потому что они имеют обратные связи и выполняют одну и ту же задачу для каждого элемента последовательности и выход зависит от предыдущих вычислений. Каждый нейрон может использовать свою внутреннюю память для хранения информации о предыдущем состоянии сети. RNN имеет обратные связи, которые позволяют передавать информацию от одного шага сети к следующему.

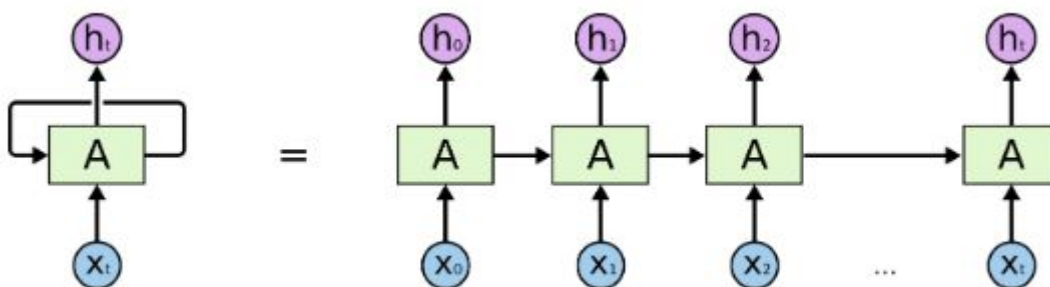


Рис. 7: Рекуррентная нейронная сеть и ее развертка во времени
[colah.github.io/posts/2015-08-Understanding-LSTMs/]

Таким образом, в RNN сети реализуется «память», что принципиально меняет характер ее работы и позволяет обнаруживать зависимости в любой последовательности данных.

Существует два вида зависимостей: краткосрочные описывают зависимости за небольшой срок, долгосрочные, с другой стороны, описывают зависимость между гораздо большим промежутком времени. Обнаружение таких зависимостей позволяет RNN распознавать шаблоны в данных и прогнозировать тенденции.

Однако фундаментальная проблема использования простых рекуррентных сетей заключается в том, что они способны обнаруживать только краткосрочные зависимости. Причиной этого является проблема отсутствия долгосрочной памяти. В простых RNN нейроны хорошо «помнят» недавно полученную информацию, но не имеют возможности надолго сохранить информацию в памяти.

Все рекуррентные нейронные сети имеют форму цепочки повторяющихся модулей нейронной сети. В стандартных RNN этот повторяющийся модуль будет иметь очень простую структуру, например он может представлять собой один слой с функцией активации \tanh (гиперболический тангенс):

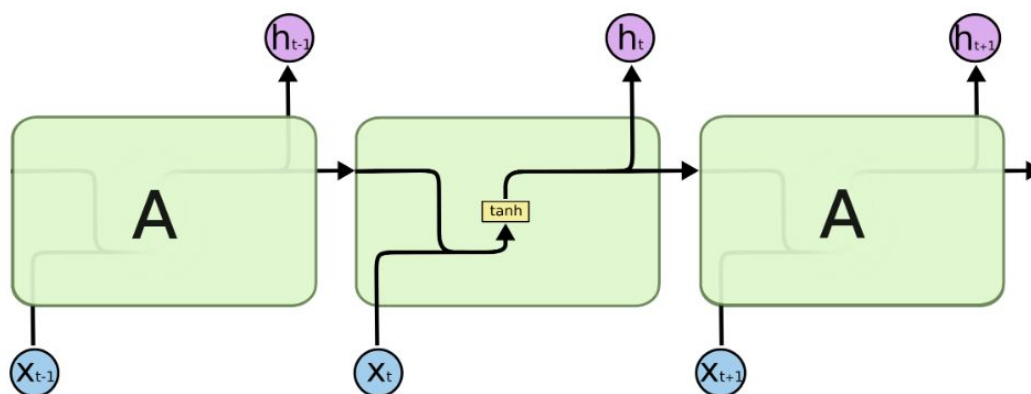


Рис. 8: Развертка RNN во времени

[colah.github.io/posts/2015-08-Understanding-LSTMs/]

Нейронные сети с долгой краткосрочной памятью (LSTM)[7] представляют собой особый тип RNN, которые имеют более длинную

«память», чем их предшественники, и способны изучать долгосрочные зависимости.

В LSTM блок работает как обычный рекуррентный блок, но имеет дополнительную ячейку памяти и различные фильтры, обычно называемые “вентили”. Данные вентили состоят из слоя сигмоидальной нейронной сети и операции поточечного умножения. Сигмоидальный слой возвращает числа в диапазоне от 0 до 1. LSTM также имеют цепочечную структуру, но повторяющийся модуль отличается. Вместо одного слоя нейронной сети существует четыре.

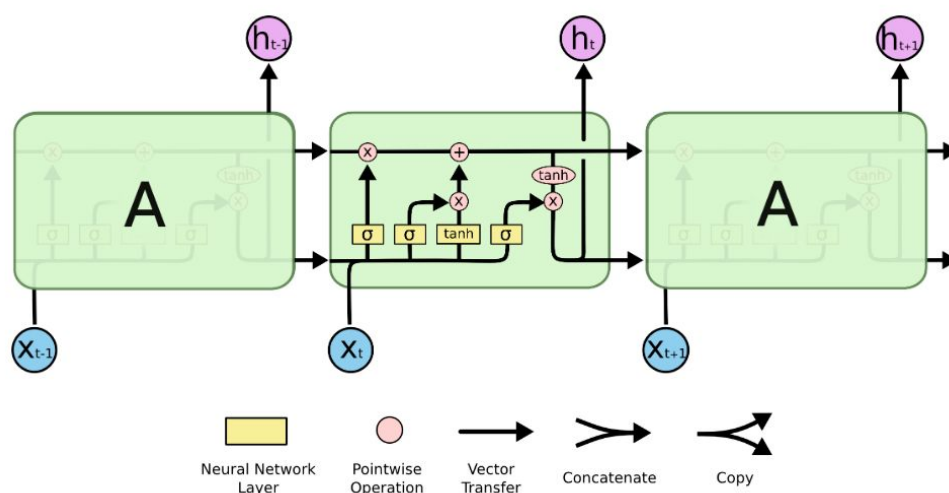
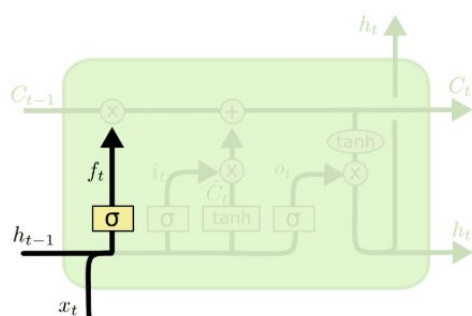


Рис. 9: LSTM сеть [colah.github.io/posts/2015-08-Understanding-LSTMs/]

Первый шаг в LSTM - решить, какую информацию выбросить из состояния ячейки, в этом нам поможет “вентиль забывания”. Этот

вентиль контролирует меру сохранения значения в памяти.

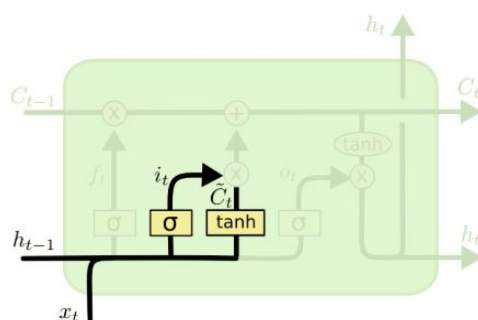


$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Рис. 10: Первый шаг работы LSTM.

[colah.github.io/posts/2015-08-Understanding-LSTMs/]

Следующий шаг - определить, какую новую информацию мы будем хранить в состоянии ячейки. Этот шаг включает в себя две части. Во-первых, «входной вентиль» решает, какие значения мы будем обновлять. Затем tanh-слой создает вектор новых значений кандидатов, которые могут быть добавлены в состояние.



$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Рис. 11: Второй шаг работы LSTM

[colah.github.io/posts/2015-08-Understanding-LSTMs/]

Далее пора обновить старое состояние ячейки на новое. Умножаем старое состояние на f_t забывая, то что решили забыть ранее, затем мы добавляем новые значения-кандидаты.

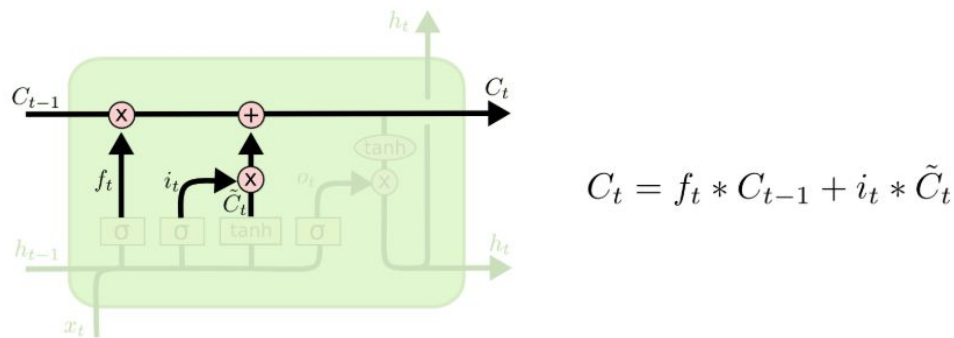


Рис. 12: Третий шаг работы LSTM

[colah.github.io/posts/2015-08-Understanding-LSTMs/]

На последнем шаге, нам нужно решить, что мы собираемся выводить. Выходные данные будут основаны на состоянии ячейки. К нему применяется «выходной клапан» контролирующей в какой степени находящееся в памяти значение, используется при расчёте выходной функции активации для блока. После значения состояния ячейки проходят через tanh-слой, чтобы получить на выходе значения из диапазона от -1 до 1, и перемножаются с выходными значениями сигмоидального слоя, что позволяет выводить только требуемую информацию.

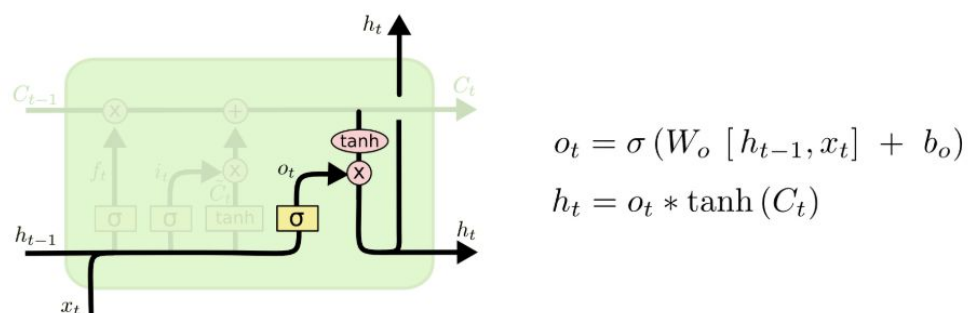


Рис. 13: Последний шаг работы LSTM

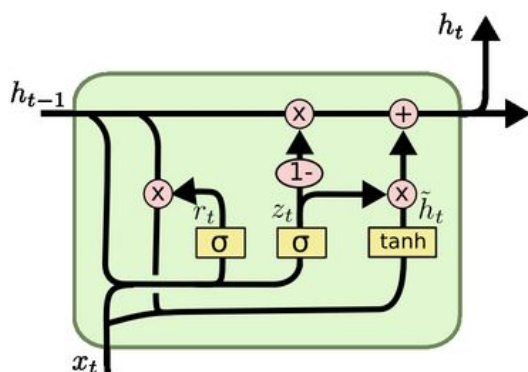
[colah.github.io/posts/2015-08-Understanding-LSTMs/]

Вентили помогают определить долгосрочные зависимости, управляя потоком данных внутри блока. Ячейка дает сети своего рода память и сохраняет долгосрочные зависимости. Процесс обучения и обратного распространения здесь работает так же как и в обычной RNN, но с добавлением, что вентили также имеют веса. Таким образом, модель становится лучше в определении правильной информации, что приводит к хорошим прогнозам.

Gated Recurrent Unit (GRU)[8] - это рекуррентная нейронная сеть, способная эффективно сохранять долгосрочные зависимости. Данная сеть была предложена только в 2014 году и её можно считать относительно новой, особенно по сравнению с LSTM, который был предложен в 1997 году.

Сети LSTM и GRU имеют дополнительные параметры, которые контролируют, когда и как обновляется их память.

Сети GRU[9] и LSTM могут фиксировать как долгосрочные, так и краткосрочные зависимости в последовательностях, но сети GRU содержат меньше параметров и, следовательно, быстрее обучаются. Сеть GRU имеет «вентиль сброса» и «вентиль обновления». «Вентиль обновления» определяет объем информации, получаемый из прошлого состояния. «Вентиль сброса» - мера сохранения введенного значения в памяти.



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Рис. 14: GRU сеть [colah.github.io/posts/2015-08-Understanding-LSTMs/]

GRU состоит из меньшего количества вентилей (GRU – 2, LSTM – 3). Это влияет на количество нелинейностей, которое приходит от входных данных, и в конечном счете, влияет на результаты вычислений. Кроме того, в GRU отсутствует ячейка памяти как и в LSTM.

2.1.3 Реализация нейронной сети

Очевидно, что в нашем сценарии обучения понадобится долговременная память. Преимуществом сетей GRU является то, что они обучаются быстрее, однако, из-за того, что архитектура сетей GRU более проста, она не позволяет сохранять данные так долго, как это может делать сеть LSTM. Поэтому сети GRU подходят для решения более простых задач, чем LSTM сети. В качестве архитектуры выберем LSTM.

Для разработки нейронной сети, будем использовать следующий стек технологий:

- Pandas - библиотека для анализа и обработки данных
- Matplotlib - библиотека для визуализации данных
- PyTorch - научный вычислительный пакет, у которого есть следующие возможности: замена NumPy для использования мощности графических процессоров, исследовательская платформа глубокого обучения, обеспечивающая максимальную гибкость и скорость.

При составлении прогноза закупок важно не допустить дефицита товара на складе. Небольшие излишки товара можно назвать страховым запасом, а отсутствие необходимого товара при обращении клиента несет, как минимум - потерю выручки, а как максимум - ущерб имиджу компании. Чтобы учесть этот фактор при обучении, в качестве функции ошибки будем использовать асимметричную MSE(1), которая обеспечивает более высокое значение ошибки, когда модель предсказывает значение ниже реального[10].

Формула асимметричной MSE:

$$\begin{cases} (y_{predict} - y_{target})^2, & y_p > y_t \\ (y_{predict} - y_{target})^4, & else \end{cases} \quad (1)$$

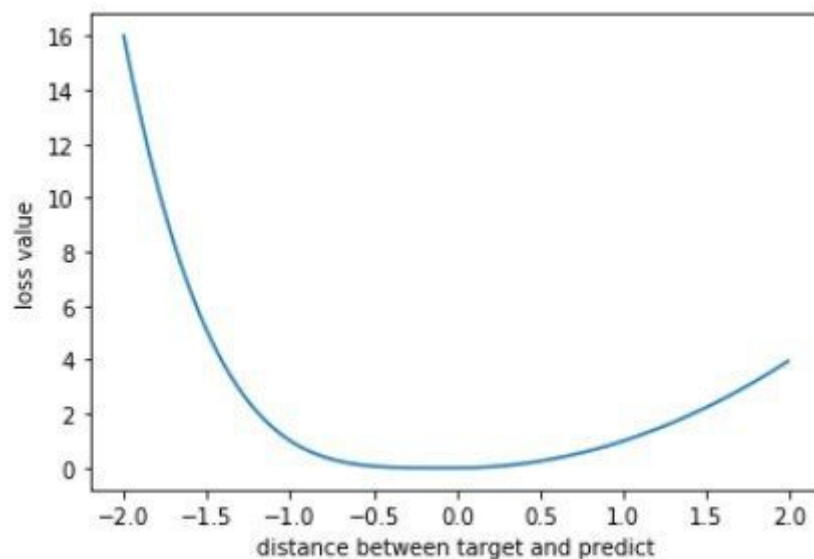


Рис. 15: Ассиметричный MSE loss

В качестве метода оптимизации будем использовать стохастический градиентный спуск с обучением на мини-батчах. В отличие от градиентного спуска, он обновляет свои веса после подсчета ошибки для каждого батча и за счет этого быстрее сходится.

Также, чтобы повысить эффективность обучения, важно уменьшать шаг обучения на каждой эпохе. В данной реализации уменьшение скорости обучения происходит раз в 10 эпох.

Чтобы не допустить переобучения нейронной сети, важно следить за ошибкой валидации. Когда она начинает возрастать, останавливаем обучение.

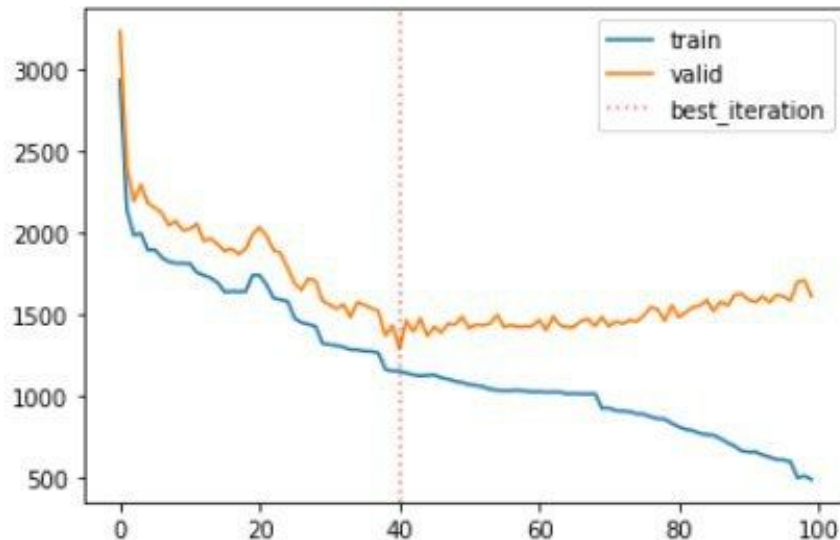


Рис. 16: Остановка обучения при возрастании ошибки валидации

Для того, чтобы проверить качество разработанной нейронной сети, будем использовать подход отложенной выборки. Выборку разобьем на две части, 80% будем использовать для обучения, а 20% данных отложим для тестирования.

```
split_idx = int(0.8 * len(df2_agg))
X_train = torch.from_numpy(
    df2_agg[['Quantity', 'is_holiday', 'lr_coef_1', 'lr_coef_2', 'nearest_hday_timedelta']].values[:split_idx]
).float()
X_test = torch.from_numpy(
    df2_agg[['Quantity', 'is_holiday', 'lr_coef_1', 'lr_coef_2', 'nearest_hday_timedelta']].values[split_idx:]
).float()
```

Рис.17: Разбиение выборки на тестовую и тренировочную

Тестирование:

	train	valid	test
--	-------	-------	------

Средняя абсолютная ошибка	33.97396061691 954	35.91977171419 663	37.12330804225 2916
Средняя квадратичная ошибка	1154.23	1290.23	1378.14
Ассиметричный MSE loss	1466.56	1722.0	1879.34

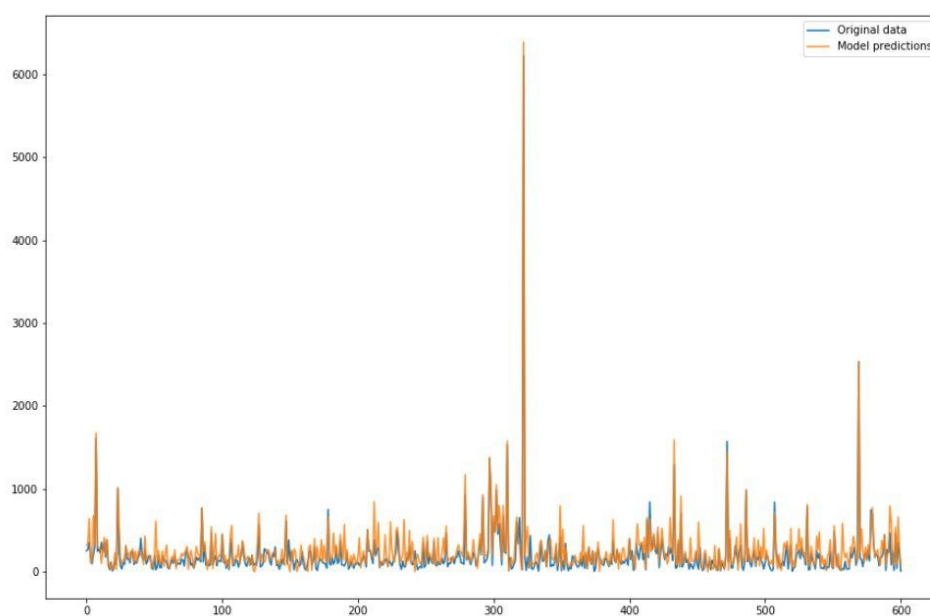


Рис.18: Предсказание для одного товара

Построим предсказания для всех товаров:

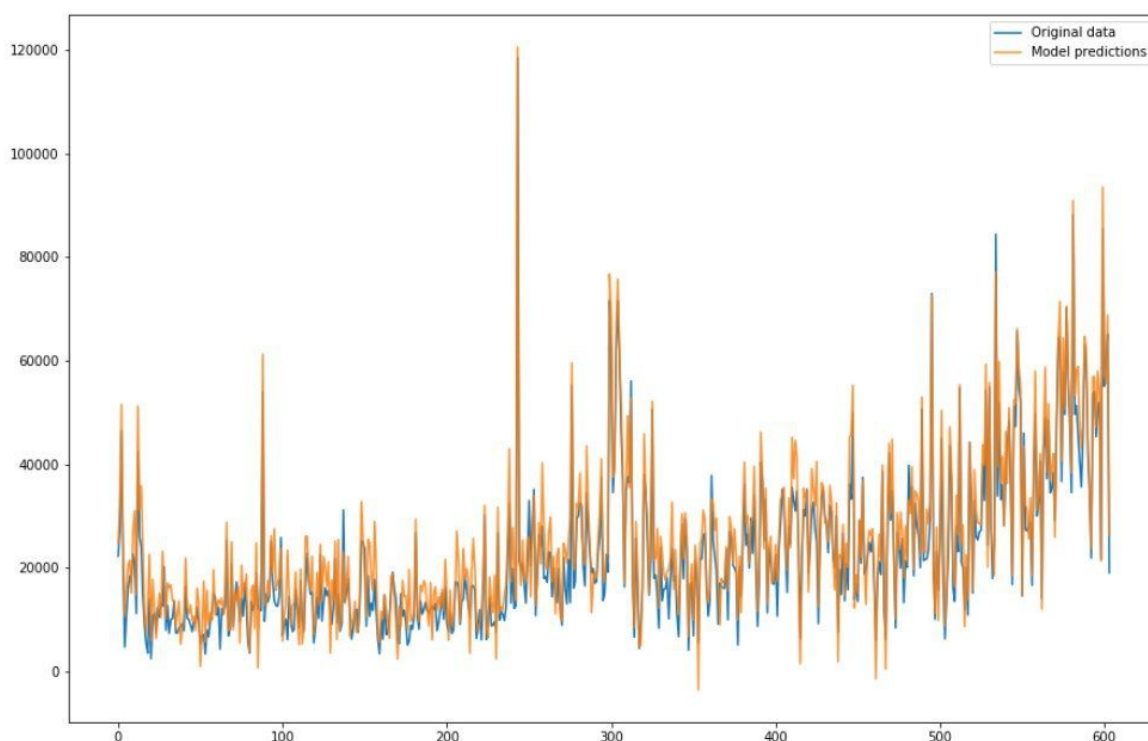


Рис.19: Предсказание для всех товаров

2.2 Разработка веб-сервиса

Для удобного взаимодействия пользователя с построенной нейронной сетью необходимо разработать интерфейс. Каждый пользователь ожидает от приложения: функциональность, скорость работы, простоту использования, безопасность и надежность.

Наиболее распространенными интерфейсами для работы с нейронными сетями являются:

- Командная строка. К преимуществам данного способа взаимодействия относится простота реализации, а из недостатков можно выделить то, что для работы с командной

строкой необходимы базовые навыки компьютерной грамотности, что может отсеять часть целевой аудитории.

- **Десктопное приложение.** Данный способ взаимодействия не имеет в себе недостатков предыдущего, но имеет ограничения на кроссплатформенность. Пользователям необходимо устанавливать приложение, что влечет за собой необходимость его разработки под разные операционные системы. Из-за этого трудозатраты на выполнение работ возрастают.
- **Мобильное приложение.** Из преимуществ можно выделить: удобство разработки, для взаимодействия с приложением пользователю не требуется специальных навыков. К недостаткам можно отнести: возможные неудобства загрузки данных; ограниченность реализации только одной платформой.
- **Веб-приложение.** Современное решение, содержащее в себе основные достоинства предыдущих, такие как: мобильность (для работы с приложением достаточно иметь доступ к браузеру), простота развертывания и обновления, масштабируемость и кроссплатформенность

В качестве способа взаимодействия с пользователем выберем веб-приложение. Веб-сервис будет содержать в себе следующие модули:

- **Клиентское веб-приложение.** Предназначено для загрузки данных, отправки REST запросов, отображения графического интерфейса и результатов прогноза пользователю.

- Модуль REST-сервиса. Предназначен для обработки запросов с клиентской стороны, передачи данных в нейронную сеть и обратно клиенту.
- Модуль нейронной сети. Предназначен для составления прогноза.

2.2.1 Реализация веб-сервиса

Для создания клиентской части веб-приложения будет использоваться платформа Angular на языке TypeScript. Angular является современным веб-фреймворком для создания SPA (single page application), использующий архитектуру MVC (Model-View-Controller). С архитектурой MVC можно изолировать логику работы приложения от пользовательского интерфейса и поддерживать разделение ответственности. Контроллер получает все запросы для приложения и работает с моделью для подготовки данных, необходимых представлению. Представление использует данные, подготовленные контроллером, и отображает окончательный вид.

Серверная часть, будет реализована на node.js. Node.js - это среда исполнения JavaScript в машинный код. Одним из несомненных преимуществ данной платформы является асинхронность в сочетании с событийным подходом. В отличие от потокового, парадигма событийно-ориентированного программирования основана на том, что выполнение программы определяется событиями - внешними действиями пользователя.

Веб-приложение должно предоставлять пользователю такой функционал как:

- Сохранение результатов предыдущих расчетов в личном кабинете.
- Загрузка данных для составления прогноза.
- Выбор временных рамок для составления прогноза.
- Получение результатов в наглядном формате и возможность скачать полученный прогноз для дальнейшей работы.

Продemonстрируем работу пользователя с разработанным веб-сервисом. После авторизации, происходит переход на главную страницу сайта, где можно загрузить данные и выбрать временной промежуток для прогнозирования.

Прогнозирование закупок Личный кабинет

Срок прогноза:

месяц

Перетащите
или
Выберите файл

OnlineRetail.csv
43.47 MB

Отправить

Рис. 20: Загрузка файла и выбор временного промежутка

После отправки пользователю возвращается результат работы в наглядном виде. Также при необходимости имеется возможность скачать результат.

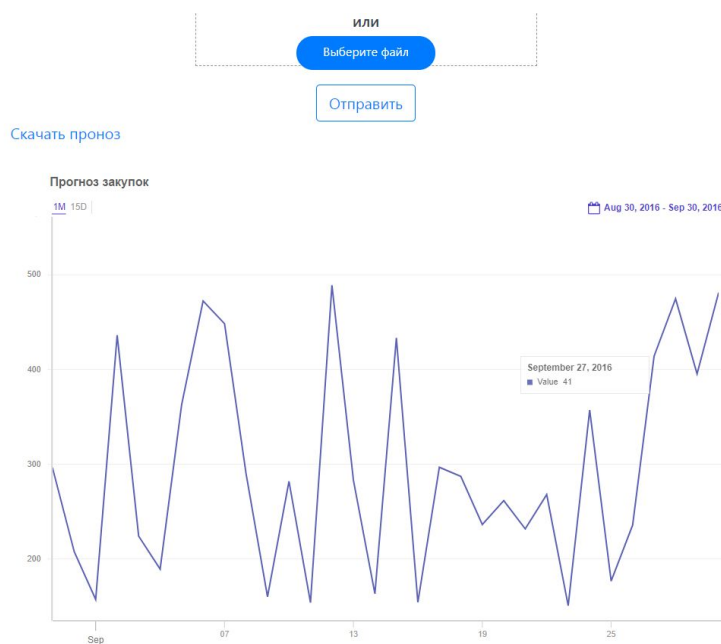


Рис. 21: Результат работы веб-приложения

Глава 3. Анализ полученного решения

Рассмотрим и проанализируем разработанное приложение, сравним с другими решениями, а также выявим основные достоинства и недостатки.

Несомненным достоинством полученного решения является то, что оно способно составлять как краткосрочные, так и среднесрочные прогнозы с хорошей точностью. При анализе уже существующих решений были выявлены такие основные недостатки:

- Приложения на базе Microsoft Excel требуют больших трудозатрат для составления прогноза. Учитывая человеческий фактор, риск совершения ошибок повышается. Созданное в ходе данной работы решение является веб-сервисом, для составления прогноза нужно просто загрузить данные и выбрать интервал прогноза.
- Статистические модели, используемые для прогнозирования, не учитывают скрытых факторов, влияющих на продажи, в отличие от них нейронные сети способны решать задачи, в которых неизвестны закономерности развития ситуации, зависимости между входными и выходными данными, а также имеются шумы — ненужная информация, которую сеть обходит.
- Крупные решения для прогнозирования необходимого количества товара, такие как Forecast Now используют в своей основе для сбора информации системы автоматизации, из-за этого они не подходят для ведения малого бизнеса.

Разработанное приложения подходит для анализа как крупных компаний, так и малого бизнеса.

Из-за большого количества подбираемых параметров нейронные сети в сравнении с линейными моделями легче переобучаются и поэтому могут давать прогнозы более низкого качества.

Для дальнейшего улучшения работы программы необходимо добавить линейные модели. Производить анализ временного ряда и выбирать подходящую модель.

Заключение

В данной работе были рассмотрены существующие решения используемые для прогнозирования закупок. В результате анализа были выявлены следующие недостатки: высокие трудозатраты для составления прогноза, отсутствие учета дополнительных факторов влияющих на продажи, привязка работы к системам автоматизации.

Были выделены основные критерии, которым должно отвечать разрабатываемое решение. Также был осуществлен анализ методов, которые можно использовать для прогнозирования, выбор архитектуры нейронной сети, подходящей для прогнозирования временных рядов. Разработана, реализована, обучена и протестирована нейронная сеть для прогнозирования закупок на заданный промежуток, учитывающая сезонность продаж.

В результате было создано веб-приложение на основе нейронной сети, позволяющее загрузить данные и получить результат в наглядном виде.

Решение было проанализировано, выявлены основные достоинства и недостатки, а также предложены возможные пути развития для дальнейшего улучшения.

Список литературы

1. Levine, David M., David F. Stephan, Timothy C. Krehbiel, and Mark L. Berenson (2011) Statistics for Managers Using Microsoft Excel. Sixth Edition. Pearson Education
2. Брошюра по Forecast PRO [Электрон. ресурс] // Forecast PRO: Обучение. Дополнительные материалы для скачивания. URL: http://www.forecastpro.ru/d/ForecastPRO_whitepages_280814.pdf (Дата обращения: 10.01.2020)
3. Moving Average (MA) — скользящая средняя [Электрон. ресурс] - URL: sweetrading.ru/indicator/moving-average-ma.html (Дата обращения: 12.05.2020)
4. Ицхоки О. Выбор модели и парадоксы прогнозирования
5. Линейная регрессия (Linear regression) [Электрон. ресурс] - URL: wiki.loginom.ru/articles/linear-regression.html (Дата обращения: 18.05.2020)
6. Сглаживание скользящих средних достоинства и недостатки [Электрон.ресурс] - URL: ebmastermaksim.ru/foreks/sgla-zhivanie-skolzyashhix-srednix-dostoinstva-nedostatki.html (дата обращения: 15.03.2020).
7. Essentials of Deep Learning : Introduction to Long Short Term Memory [Электрон.ресурс] - URL: www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/ (дата обращения: 16.03.2020).

8. Understanding GRU Networks [Электрон.ресурс] - URL: towardsdatascience.com/understanding-gru-networks-2ef37df6c9be

9. Recurrent Neural Network Tutorial, Part 4 – Implementing a GRU/LSTM RNN with Python [Электрон.ресурс] URL - www.wildml.com/2015/10/recurrent-neural-network-tutorial-part-4-implementing-a-grulstm-rnn-with-python-and-theano/ (дата обращения 12.05.2020)

10. Kevin Aretz, Söhnke M, «Асимметричные функции потерь и рациональность ожидаемой доходности фонда». Международный журнал по прогнозированию.